

The Battle of Neighborhoods

1. Introduction:

The goal of this project is to help people explore the best places in their neighborhood. This will help them to make smart and efficient decisions about choosing a good neighborhood among other neighborhoods in North York, Toronto.

Many people migrate to various states in Canada and need a lot of research to get good housing prices and reputable schools for their children. This project is for people looking for better neighborhoods. To facilitate access to cafe, school, super market, medical stores, grocery stores, mall, theater, hospital, like-minded people, etc.

This project aims to create an analysis of the characteristics for a people migrating to North York to find a better neighborhood as a comparative analysis between neighborhoods. Features include the median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

It will help people become aware of the area and neighborhood before moving to a new city, state, country or place for work or to start a new life

Problem to solve:

The main goal of this project is to suggest a better neighborhood in a new city for the person who moves there. Social presence in society in terms of like-minded people. Connectivity to airport, bus stop, city center, markets and other daily needs nearby.

1. Sorted list of houses in terms of housing prices in ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

Location:

North York is one of the six administrative districts of Toronto, Ontario, Canada. It was one of the fastest-growing parts of the region due to its proximity to Old Toronto. It was declared a borough in 1967, and later became a city in 1979, attracting high-density residences, rapid transit, and a number of corporate headquarters in North York City Centre, its central business district. In 1998, North York was amalgamated with the rest of Metropolitan Toronto to form the new city of Toronto and has since been a secondary economic hub of the city outside Downtown Toronto.

Foursquare API:

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

2. Data Description

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Will use North York dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

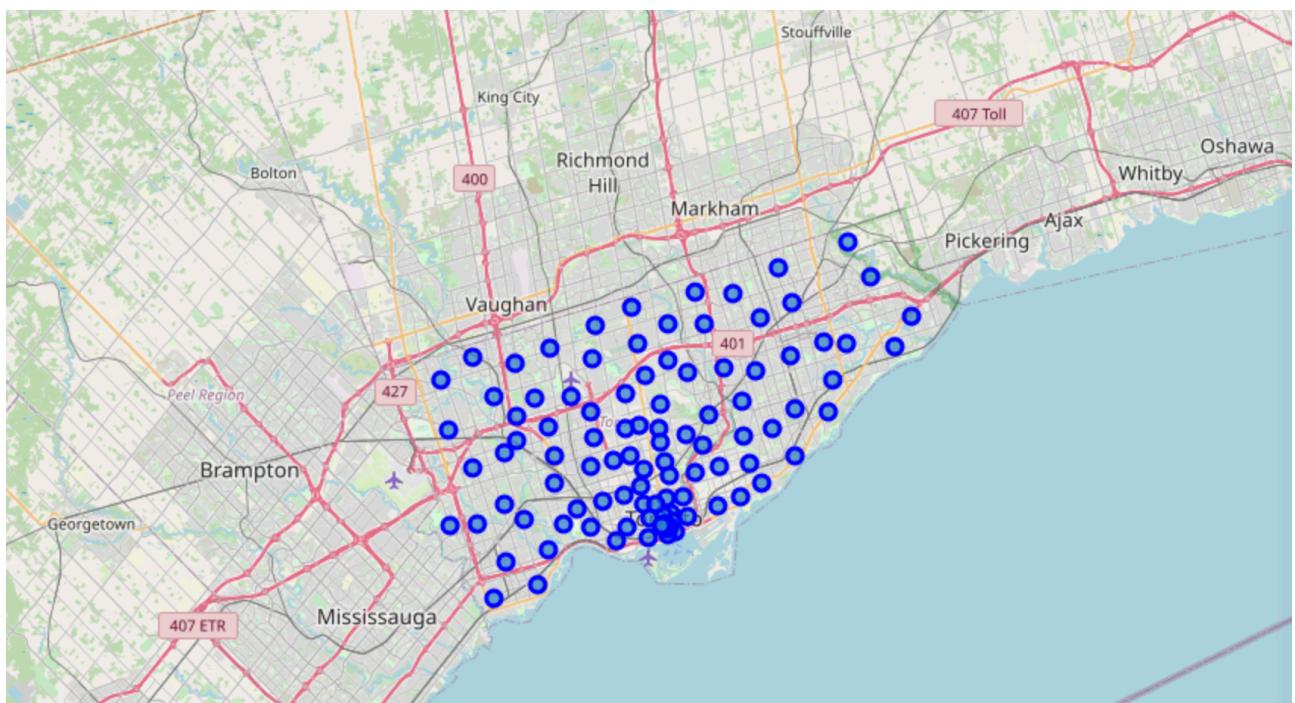
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare"

locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter. The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Map of North York, Toronto



3. Methodology Section

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach

```
Entrée [44]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

NorthYork_merged = df_2.iloc[17:35,:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
NorthYork_merged = NorthYork_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

#NorthYork_merged.head(18) # check the last columns!
NorthYork_merged
```

Out[44]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
17	M2H	North York	Hillcrest Village	43.80225	-79.35558	0	Park	Bakery	Fast Food Restaurant	Pharmacy	Chinese Restaurant	Residential Building (Apartment / Condo)	Yoga Studio
18	M2J	North York	Fairview, Henry Farm, Oriole	43.78097	-79.34781	1	Clothing Store	Coffee Shop	Bakery	Fast Food Restaurant	Sandwich Place	Bank	Juice Bar
19	M2K	North York	Bayview Village	43.78112	-79.38060	0	Trail	Flower Shop	Gas Station	Asian Restaurant	Park	Elementary School	Dry Cleaner
20	M2L	North York	York Mills, Silver Hills	43.75698	-79.38060	0	Concert Hall	Park	Event Space	Dry Cleaner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store

Most Common venues near Neighborhood

```
Entrée [42]: import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}_th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = NorthYork_grouped['Neighborhood']

for ind in np.arange(NorthYork_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(NorthYork_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[42]:

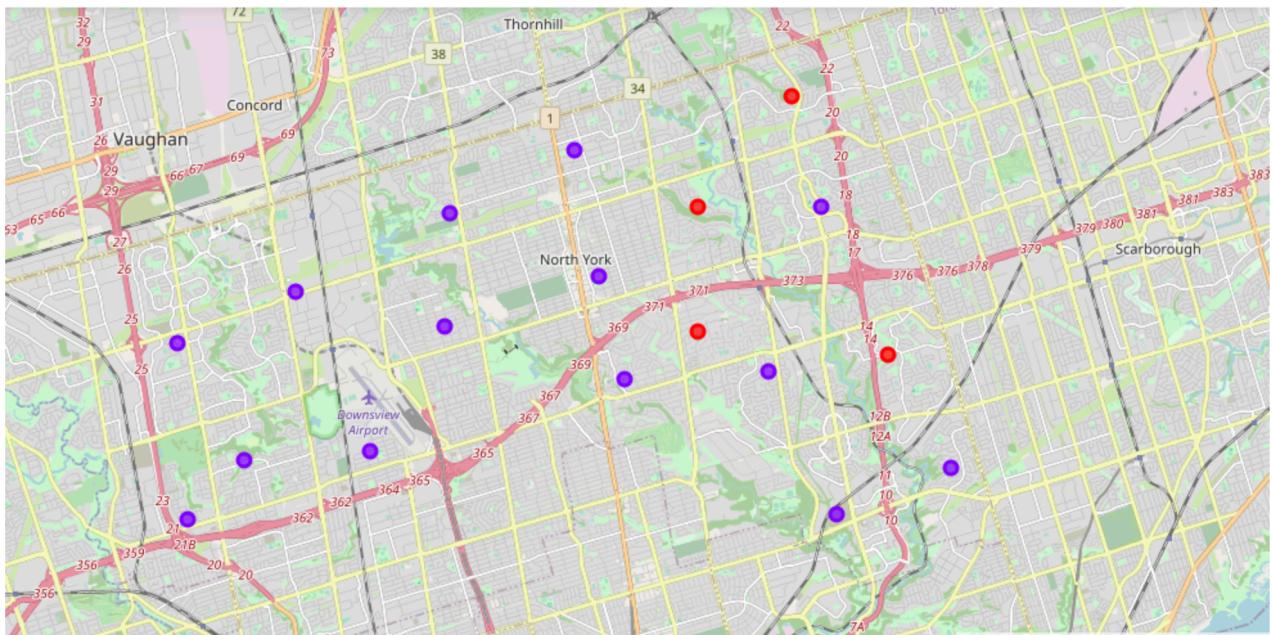
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aigincourt	Shopping Mall	Chinese Restaurant	Pharmacy	Coffee Shop	Breakfast Spot	Café	Latin American Restaurant	Sandwich Place	Supermarket	Sushi Restaurant
1	Alderwood, Long Branch	Sandwich Place	Pub	Coffee Shop	Pharmacy	Gym	Gas Station	Pizza Place	Eastern European Restaurant	Dive Bar	Dog Run
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Coffee Shop	Grocery Store	Fried Chicken Joint	Sandwich Place	Mediterranean Restaurant	Sushi Restaurant	Deli / Bodega	Restaurant	Pizza Place
3	Bayview Village	Trail	Flower Shop	Gas Station	Asian Restaurant	Park	Elementary School	Dry Cleaner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Sandwich Place	Boutique	Butcher	Juice Bar	Sushi Restaurant	Sports Club	Liquor Store	Pharmacy

Work Flow:

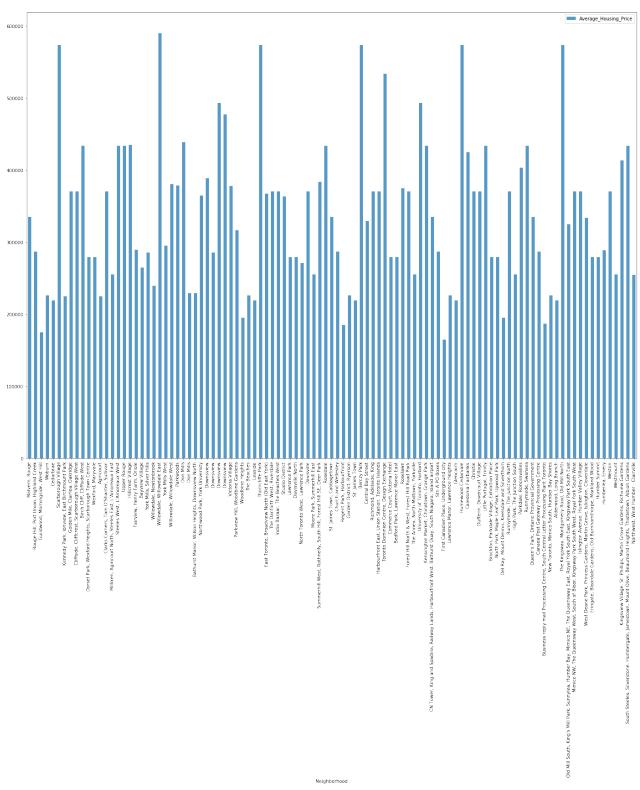
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

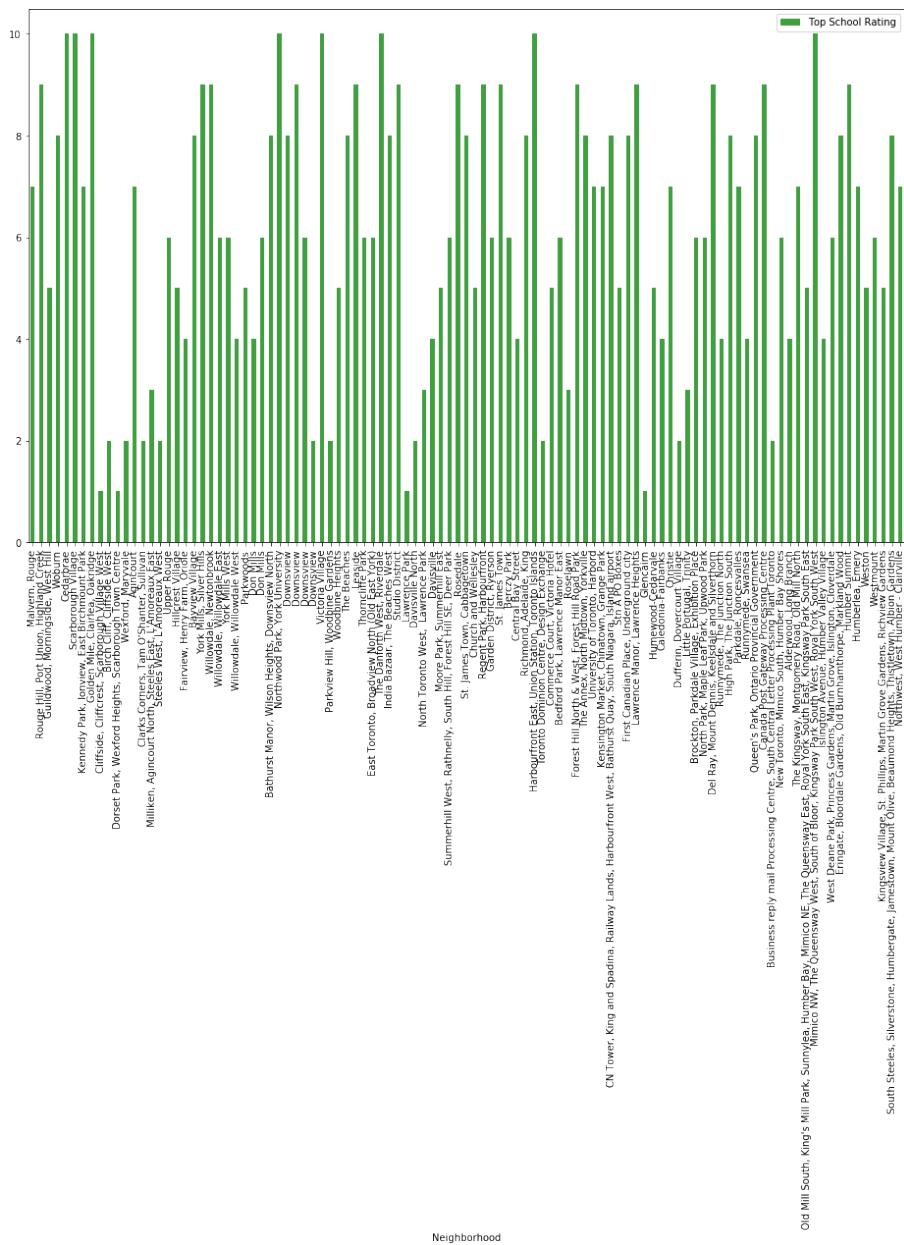
Map of Clusters in North York



Average Housing Price by Clusters in North York



School Ratings by Clusters in North York



The Location:

North York is one of the six administrative districts of Toronto, Ontario, Canada. It was one of the fastest-growing parts of the region due to its proximity to Old Toronto. It was declared a borough in 1967, and later became a city in 1979, attracting high-density residences, rapid transit, and a number of corporate headquarters in North York City Centre, its central business district. In 1998, North York was amalgamated with the rest of Metropolitan Toronto to form the new city of Toronto and has since been a secondary economic hub of the city outside Downtown Toronto.

Foursquare API:

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

Problem to Solve:

The main goal of this project is to suggest a better neighborhood in a new city for the person who moves there. Social presence in society in terms of like-minded people. Connectivity to airport, bus stop, city center, markets and other daily needs nearby.

1. Sorted list of house in terms of housing prices in a ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

6. Conclusion

In this project, using the k-means cluster algorithm, I split the neighborhood into 10 (ten) different clusters and for 103 different latitudes and longitudes of the dataset, which have very neighborhoods. similar around them. Using the graphs above, the results presented to a particular neighborhood based on the average house prices and the school rating were obtained.

I feel rewarded for the effort and I think this course, with all the topics covered, is worth enjoying. This project showed me a practical application for solving a real situation with personal and financial impact using Data Science tools.

Mapping with Folium is a very powerful technique to consolidate information and improve analysis and decision with confidence.

Future Works:

This project can be continued to make it more specific in terms of finding the best home in North York. Best way based on all the things required (daily needs or things we need to live a better life) and also in terms of profitability.